Ed Guinness, Susie Slavney, and Tom Stein
PDS Geosciences Node
PDS MC, Pasadena, Nov. 18-19, 2013

InSight Archive Development Report

InSight Archive Development Status

DAWG is up and running with monthly telecons.
Currently, PDS nodes are having initial meetings with
instrument teams to begin data product design.

- HP3/RAD/Geo: Teleconference held November 12.
- IDC/ICC/IDA/Imaging/Geo/MIPL: Meeting at JPL Wednesday, November 20.
- RISE/Geo: May also meet with RISE P.I. on Wednesday.
- SEIS/Geo: Teleconference scheduled Friday, November 22.
- APSS/Atm: Meeting to be scheduled with team.
- Magnetometer/PPI: Products will be binary tables, similar to Cassini products from the same team. More discussion to come.

InSight Archive Development Status

Working on an archive development schedule

- We're trying to mesh EDR development with MIPL software deliveries, but MIPL's schedule is uncertain. To be discussed at JPL meeting on Wednesday.
- Mission System CDR is September 17, 2014, but it's not clear what level of archive development should be done by then.
- Proposed due dates, as of last DAWG, October 30:
  - **January 31, 2014**: All raw data products defined (e.g.., sample products available, labels designed, drafts of documentation available).
  - **September 4, 2014**: All derived products defined.
  - These dates to be revised after individual discussions with teams. At present it appears the January date will be too early for MIPL.

Use of SEED Format for Seismic Data

- SEIS is the seismometer on the InSight lander. The instrument team is based at IPGP, the Paris Institute for Earth Physics.
- SEIS data are compressed in SEED format on the spacecraft and downlinked that way.
- The SEIS team and the InSight P.I. want to archive the raw data in their original condition.
- Higher-level SEIS products are still being defined, and the team has not decided which, if any, should use SEED format.

Use of SEED Format for Seismic Data

SEED = Standard for the Exchange of Earthquake Data

- International standard since 1987 now in wide use
- Specification:
  www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf
- Allows lossless compression using a Steim algorithm (basically first-difference compression)
- Software to convert between SEED and other formats, including ASCII, is freely available. Here is one list:
  www.iris.edu/dms/nodes/dmc/software/downloads/

Issues

1. Can raw SEIS data in SEED format be archived in PDS?
2. Can derived SEIS data in SEED format be archived in PDS?
3. How can raw SEIS data be described using the PDS4 standard?
4. Why not just allow SEIS data to go to other archiving agencies, and thereby avoid dealing with it?

Issue 1. Can raw SEIS data in SEED format be archived in PDS?

- As raw SEIS data will be compressed on the spacecraft and downlinked in SEED format, the PDS policy regarding compressed data applies.
- SEIS data compressed in SEED format meet the conditions of the policy. (See backup slides for complete text of policy.)

    - Data are received from the spacecraft in compressed form.
    - The decompression algorithm is non-proprietary.
    - The decompression algorithm can be provided in the archive.
    - Source code for decompression can be provided in the archive.
    - Examples of data before and after decompression can be provided in the archive.

Issue 2. Can derived SEIS data in SEED format be archived in PDS?

- It's uncertain whether the SEIS team wants to use SEED for derived data products. This will be discussed in a teleconference on Nov. 22.
- Let's postpone this discussion until we know whether the team will request to use compressed SEED for derived products.

Issue 3. How can raw SEIS data be described using the PDS4 standard?

- Encoded byte stream?
  - Would require change to the information model to allow encoded byte stream to be used for observational products.

Issue 4. Why not just allow SEIS data to go to other archiving agencies?

- SEIS may be able to satisfy its archiving requirement by delivering to Geoscope and IRIS, without delivering to PDS.
- Bruce Banerdt, InSight P.I.: "I am still committed to delivering all InSight data to the PDS, including SEED-y seismic data, regardless of any legal loopholes that might be available to me."
- The PDS archive for InSight would be conspicuously incomplete without data from SEIS, one of the two primary instruments. This would not help PDS' reputation and might set an undesirable precedent. ("Hmm, PDS is too hard; who else can we send our data to?")

Next Steps

- Proceed to design PDS4 labels for raw SEIS data in SEED format.
- Better understand derived products and whether SEED is a desirable format for them.

Backup Slides

- Policy for Use of Compression in PDS Archives
- SEED Format Illustration

Policy for Use of Compression in PDS Archives, page 1 of 3

Data may not be archived in compressed form except as follows.

1. Large volumes of data received from spacecraft in compressed form may be archived in that compressed form subject to the following conditions:

   a. The decompression algorithm must be non-proprietary.

   b. A detailed decompression algorithm (or reference to a detailed algorithm in published literature) must be provided as part of the archive.

   c. Software source code for decompression must be provided in at least one high-level programming language in common use by the science community (e.g., C, C++, Java, Fortran, or IDL). The source code is intended not as operational software but as a "skeleton" program that can be adapted to new computer systems and operating environments. The source code captures the subtle implementation considerations of the compression algorithm that may not be apparent from the algorithm description.

   d. Before and after examples of the data decompression algorithm must exist in the archive for testing software implementations of the decompression algorithm.

(pds-engineering.jpl.nasa.gov/charter_policy/compression_policy_20051114.htm)

2. With explicit permission from the appropriate Discipline Node, derived image products may be archived in compressed form under the following conditions:

   a. The source version of the image products must be archived in uncompressed form.
   b. Compression must be lossless unless the lossless requirement is explicitly waived by the Discipline Node.
   c. The compression format must be approved in advance by the Discipline Node and the Management Council.
   d. Product meta-data must identify the compression algorithm (or software) used in compressing the product, and its version
   e. All PDS meta-data for each product must be available in uncompressed form
   f. The PDS must have a copy of the specification or standard defining the compression algorithm used, at the version level that was used. If legally permitted, the documentation should be included in the archive.
   g. Decompression software must be capable of producing a correctly formatted and labeled decompressed PDS data file. Additional output formats are permitted. Source code and executables for decompression programs must be provided to the appropriate PDS Discipline Node at the time an archive is delivered. Well documented decompression algorithms must be included in the archive
   h. The compression and decompression software must be validated on a number of test data files to verify that the input and output files are identical. Thereafter, a random sampling of data products in the archive should be decompressed as part of the validation process.
   i. The compressed products must be validated to comply with the specification or standard defining the compression algorithm used.
   j. The compressed products, decompression algorithms, and decompression software must all be available for use by the PDS and its users on a royalty and license fee free basis

3. Compression of other files is allowed subject to the following conditions:
   a. Lossless compression software from INFOZIP will be used; a PDS minimal label with pointer to INFOZIP will accompany the compressed file
   b. PDS will capture the INFOZIP software tree at least annually and make it available for distribution
   c. Files critical to understanding structure and basic content of the archive will NOT be compressed

Each Discipline Node accepting compressed data must keep an inventory of those holdings and take action to maintain the usability of the data as needed.

SEED Format Illustration

SEED Version 2.1 and later Style

| Volume Index Control Header |
| Abbreviation Dictionary Control Header |
| Station Control Headers |
| Time Span Control Header 1 |
| Time Span Control Header 2 |

.
.
.

| Time Span Control header N |
| Data for Time Span 1 |
| Data for Time Span 2 |

.
.
.

| Data for Time Span N |

- Header and data records can be stored in one file or
- Headers can be in a separate file (Dataless SEED)
- And data in a separate file (miniSEED)

| Identification Block |
| Fixed Header Section |
Variable Header Section	Blockette 1
	.
	Blockette N
Data Section